DOCUMENT RESUME

ED 039 092 RE 002 437

AUTHOR Johnson, Dale D.
TITLE Vowel Cluster-Phoneme Correspondences in 20,000 English Words.
INSTITUTION Wisconsin Univ., Madison.
PUB DATE 69
NOTE 5p.

EDRS PRICE EDRS Price MF-$0.25 HC-$0.35
DESCRIPTORS *Phonemes, *Pronunciation, *Spelling, *Vowels, Word Frequency, *Word Lists

ABSTRACT

The symbol-sound correspondence status of vowel-cluster (two or more adjacent vowel letters) spelling in American English was investigated. The source of the study was Venezky's 1963 revision of the Thorndike Frequency Count. A computer print-out of the 20,000 word corpus was analyzed to determine the letter-sound correspondence of vowel cluster spelling. Totals and pronunciations for each correspondence, as well as representative word lists were compiled. The analysis revealed (1) There are 61 vowel clusters representing 92 different single vowel phonemes and phoneme strings, producing more than 300 symbol-sound correspondences. (2) There was great variance in the frequency of the 61 vowel clusters. (3) Vowel clusters vary greatly in the number of individual phonemes or phoneme strings they represent. (4) Most vowel cluster pronunciations are unpredictable from their spelling. (WB)

ED039092

A B S T R A C T

VOWEL CLUSTER - PHONEME CORRESPONDENCES in 20,000 ENGLISH WORDS

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

By Dale D. Johnson
Assistant Professor
Elementary Education
Ball State University
Muncie, Indiana

Member International Reading Assoc.
National Council of Teachers of English
American Educational Research Assoc.
Phi Delta Kappa

The research which is reported below was conducted at the University of Wisconsin in 1969 with the guidance of Thomas Barrett and Richard Venezky.

This study was designed to answer the question: "What is the symbol-sound correspondence status of vowel cluster spelling in American English?" The source of the study was a corpus of 20,000 common English words - a 1963 Venezky revision of the Thorndike Frequency Count (1941). A computer program was developed by Venezky which derived and tabulated all letter-sound correspondences within the corpus.

Venezky's unpublished computer print-out was analyzed by the investigator, to determine the letter-sound correspondence of vowel cluster spellings, (2 or more adjacent vowel letters). Totals and pronunciations for each correspondence, as well as representative word list were compiled. This analysis disclosed, among other things, the following:

1. There are 61 vowel clusters (including those containing the semi-vowels w and y) in the corpus.
2. These 61 vowel clusters represent 92 different vowel phonemes and phoneme strings, producing more than 300 symbol-sound correspondences.

RE002 437

3. About one-third of the words in the corpus contain vowel clusters.

4. There is great variance in the frequency of the 61 vowel clusters. One, io, occurs in more than 1,000 words, while 26, ii, eue, etc., occur in only one word.

5. Vowel clusters vary greatly in the number of individual phonemes or phoneme strings they represent: some represent only one sound while one, ea, represents 17 sounds.

6. Vowel clusters represent one phoneme about 80% of the time.

7. Of the 26 most common vowel clusters, those occurring in 50 or more words, only 4 follow the "first vowel long, second vowel silent" generalization in 75% or more of their occurrences - ai, ay, ee, and oa. Two more, ea and ow, adhere to this slightly more than 50% of the time. With the remaining 20, the generalization is seldom or never true.

8. Most vowel cluster pronunciations are unpredictable from their spellings.

VOWEL CLUSTER - PHONEME CORRESPONDENCES

IN 20,000 ENGLISH WORDS

By Dale D. Johnson
Assistant Professor
Elementary Education
Ball State University
Muncie, Indiana

Member International Reading Assoc.
National Council of Teachers
of English
American Educational
Research Assoc.
Phi Delta Kappa

The research which is reported below was conducted at the University of Wisconsin in 1969 with the guidance of Thomas Barrett and Richard Venezky.

Reading includes the translation from spelling to sound, and the vowel clusters (2 or more adjacent vowel letters) are perhaps the most complex and unpredictable components of the letter-sound correspondence code. Vowel cluster spellings differ from single vowel spellings in several ways. They rarely appear before geminate consonant clusters; some, such as ai and au, occur infrequently in word final position, while others, such as oa and ie, rarely begin a word in English.

Some vowel clusters have a major phonemic correspondent, and several minor correspondents. For example, the major correspondent of ai is /e/ as in bait, and represents this sound 85% of the time that it occurs. It represents /ə/, villain; /æ/, aisle; /ɛ/, again; /ɑ/, plaid, and others much less frequently. Other vowel clusters have two or more major correspondents, as well as minor correspondents. The vowel cluster ow is /o/ as in own 51% of the time and /au/ as in owl 48%. The only minor correspondent is /a/ as in knowledge.

By contrast, all single vowel spellings have two major correspondences (e.g., a is /e/ or /æ/ as in rate and rat) plus several minor correspondences.

As part of an interdisciplinary study of the reading process begun at Cornell University in 1961, Venezky developed a computer program to derive and tabulate letter-sound correspondences as in a corpus of 20,000 common English words (1963). The 20,000 word corpus was a modification of the most common 20,000 words according to the Thorndike Frequency Count (1941). Venezky omitted many archaic and low-frequency words, particularly proper nouns, and added a number of words in their place. Along with other information, the computer analysis provided an inclusive tabulation of letter-sound correspondences found in the corpus as well as totals and percentages for each pronunciation in each word position, and a complete word list for each correspondence. A Pronouncing Dictionary of American English (Kenyon and Knott, 1953) was used to determine the pronunciation of most words in the corpus.

Venezky's unpublished computer print-out of spelling-to-sound correspondences in 20,000 English words was analyzed by this writer to determine letter-sound correspondences for contiguous vowels. This analysis disclosed the following:

1. There are 61 vowel clusters (including those containing the semi-vowels w and y) in the corpus.
2. These 61 vowel clusters represent 92 different single vowel phonemes and phoneme strings producing more than 300 symbol-sound correspondences.

For example, oa represents /o/, /ɔ/, /oa/ and other phonemes and phoneme strings. Yet each of these and others, are represented by a variety of spellings. Consequently, there are over 300 symbol-sound correspondences.

3. These 61 vowel clusters appear more than 6,000 times in 20,000 word corpus.

4. There is great variance in the frequency of the 61 vowel clusters. One occurs in more than 1,000 words while 26 occur in three words or less.

5. Vowel clusters vary greatly in the number of individual phonemes or phoneme strings they represent; some represent only one sound while one represents 17 sounds.

6. Most vowel cluster pronunciations are unpredictable from their spellings.

7. Of the 61 vowel clusters, 30 occur in 10 or more words in the corpus. Of these 30, 23 occur in words in which the vowel cluster is sometimes disyllabic. Only six of these vowel clusters are disyllabic more often than monosyllabic. Thus, these 30 vowel clusters, occurring in more than 6,000 words, represent single vowel phonemes about 80% of the time and two or more phonemes about 20%.

If accepted, the presentation at IRA would include the four most common pronunciations of each of the 26 vowel clusters which occur in 50 or more words, together with the percentages and number of words for each vowel cluster-sound correspondence.